

## 竞争风险模型在女性乳腺癌预后预测研究中的应用

胡海兰<sup>1</sup>, 袁满琼<sup>2</sup>, 方 亚<sup>2</sup> (厦门大学 1. 经济学院计划统计系; 2. 公共卫生学院, 福建厦门 361005)

**摘要:**目的 探索竞争风险模型在乳腺癌患者预后预测中的应用, 并与传统生存分析结果进行比较。方法 运用K-M方法估算总体风险率; 采用原因别风险模型进行多因素分析, 评价复发和转移的影响因素。结果 K-M方法的风险率高于竞争风险情况下的风险率。肿瘤大小和个人乳腺疾病史是复发的独立预后因素。结论 竞争风险模型在乳腺癌患者的预后预测中更合理、更客观。

**关键词:** 乳腺癌; 竞争风险模型; 原因别风险模型

中图分类号: R 737.9

文献标识码: A

文章编号: 1005-4057(2016)01-0009-04

DOI: 10.3969/j.issn.1005-4057.2016.01.003

### Application of competing risk model in the prognostic prediction study of female breast cancer

HU Hai-lan<sup>1</sup>, YUAN Man-qiong<sup>2</sup>, FANG Ya<sup>2</sup> (1. Department of Planning & Statistics; 2. School of Public Health, Xiamen University, Xiamen 361005, China)

**Abstract:** Objective To explore the application of competing risk model in the prognostic prediction study of female breast cancer, and compare it with traditional survival analysis method. Methods K-M method, in the present of competing risk factors, was used to estimate overall risk ratio; cause-specific risk model was used for multiple factor analysis and to evaluate the influence factors of recurrence and metastasis. SPSS16.0, excel 2007 and R2.10.1 were used in the study. Results The risk ratio estimated with K-M method was higher than that with competing risk model. Conclusion The results showed that competing risk model was more rational than the traditional survival analysis method.

**Key words:** Breast cancer; competing risk model; cause-specific risk model

乳腺癌是女性最常见的恶性肿瘤之一, 严重影响妇女的身心健康。据资料统计, 我国每年女性乳腺癌发病16.9万, 是女性第2位常见恶性肿瘤, 4.5万女性因乳腺癌死亡, 是女性第6位常见恶性肿瘤死亡原因<sup>[1]</sup>。如何更客观预测乳腺癌患者的预后及其评价影响因素显得尤为重要。生存分析是预后研究中一种比较常见的统计分析方法, 经典的生存分析一般只考虑一个终点事件, 而竞争风险模型考虑的不止是一个终点事件, 且对任一病例来讲, 其中任何一个终点事件的发生会影响其它事件的发生<sup>[2]</sup>, 那么, 这两种方法对乳腺癌患者的预后预测有何区别, 哪种方法更为科学客观呢? 本文将探讨竞争风险模型在乳腺癌患者预后预测中的应用, 并与传统生存分析结果进行比较。

### 1 资料和方法

基金项目: 国家自然科学基金(No.30671822、No.81573257)

收稿日期: 2015-11-16; 修订日期: 2015-12-26

作者简介: 胡海兰(1986-), 女, 在读研究生。

通信作者: 方 亚, E-mail: fangya@xmu.edu.cn。

#### 1.1 资料来源

研究数据来自福建省厦门市4家医院1998-2008年入院接受手术治疗的乳腺癌患者共246例, 随访结束时间为2009年5月。所有患者均为女性。所有数据均由实际问卷调查获得。以手术作为本次研究的起始事件, 相应的手术时间即为观察开始的时间。由于直接死亡的病例数太少, 所以本文将术后复发和转移视为竞争风险事件。本文所有数据的分析方法均采用软件Excel2007和SPSS16.0以及R2.10.1完成。

#### 1.2 竞争风险模型

为了进行原因别的比例风险回归, 考虑协变量 $X$ , 并且发生终点事件 $k$ 的原因别风险模型如下:

$$\lambda_k(t) = \lambda_{k0}(t) \exp(\beta_k^T X) \quad (1)$$

式(1)中,  $\lambda_{k0}(t)$ 表示终点事件 $K$ 的原因别基础风险,  $\beta_k$ 表示对终点事件 $K$ 的回归系数, 对于不同的终点事件有不同的系数 $\beta_k$ 。

生存函数是指没有发生任何终点事件的概率。表示为:

$$S(t) = \exp\left(-\sum_{1 \leq k \leq K} \int_0^t \lambda_k(s) ds\right) \quad (2)$$

总累积风险函数为

$$F(t)=P(T \leq t)=\sum_{1 \leq k \leq K} \int_0^t \lambda_k(s) S(s) ds \quad (3)$$

### 1.3 非参估计方法

我们取  $0 < t_1 < t_2 < \dots < t_N$  代表各种终点事件发生的时间点, 用  $d_{kj}$  表示在时间  $t_j$  发生终点事件的病例数, 且用  $d_j = \sum_{k=1}^K d_{kj}$  代表在时间发生各种终点事件的病例数的总和。用  $n_j$  表示在危险集中的病例数, 即指那些在时刻  $t_j$  仍然在观察集中并且没有发生任何终点事件的病例数。

离散的原因别风险公式为:

$$\lambda_k(t) = \text{Prob}(T = t_j, D = k | T > t_{j-1}) \quad (4)$$

估计为:

$$\hat{\lambda}_k(t) = \frac{d_{kj}}{n_j} \quad (5)$$

在时刻  $t_j$  发生终点事件  $K$  的非条件概率为:  $P_k(t_j) = \text{Prob}(T = t_j, D = k)$ , 可以表示为:  $P_k(t_j) = \lambda_k(t_j) \cdot P(T > t_{j-1})$ , 即为瞬间风险率与时刻  $t_j$  的生存率的乘积。

最后, 终点事件  $K$  在时刻  $t$  的累积风险率可以估计为:

$$\hat{I}_k(t) = \sum_{j: t_j \leq t} \hat{P}_k(t_j), \quad \hat{P}_k(t_j) = \hat{\lambda}_k(t_j) \hat{S}(t_{j-1}), \quad \hat{\lambda}_k(t_j) = \frac{d_{kj}}{n_j} \quad (6)$$

## 2 结果

246例乳腺癌患者中直接死亡的有6例, 以复发、转移为第一终点事件的分别有39、33例, 在随访结束时仍然存活或删失的有168例。其中以复发作为第一终点事件的患者复发后存活的有24例, 复发后死亡的有15例; 以转移作为第一终点事件的患者转移后存活的有10例, 转移后死亡的有23例。所以本文将复发和转移看作互为竞争风险事件。

本研究的随访终止日期为2009年5月, 每个观察对象的随访时间以月为单位记录, 最短为1月, 最长为159月, 中位随访时间为29.14月。资料中手术年龄最小的为21岁, 最大值为86岁, 中位数为48.35岁。

### 2.1 原因别比例风险模型的结果

单因素分析结果显示, 术前淋巴结状况、临床分期、肿瘤大小、手术方式、治疗方式、个人乳腺疾病史、家族乳腺疾病史是影响复发的预后因素。在不考虑竞争风险的情况下采用Cox模型逐步回归

进行多因素分析, 得到术前淋巴结、肿瘤大小、个人乳腺疾病史是影响乳腺癌的独立预后因素。

下面在考虑竞争风险的情况下, 分析术前淋巴结(lymph)、肿瘤大小(tumour)、个人乳腺疾病史(history)对复发和转移的影响。R程序如下:

```
>coxph(Surv(time,status==1)~lymph+tumour+history,data=si)
```

```
>coxph(Surv(time,status==2)~lymph+tumour+history,data=si)
```

得到的结果如表1、2所示。

表1 复发的系数回归结果

项目	Coef	exp(coef)	se(coef)	Z	P值
淋巴结	0.230	1.258	0.321	0.716	4.7e-01
肿瘤大小	0.860	2.362	0.265	3.249	1.2e-03
个人乳腺疾病史	1.512	0.221	0.330	-4.577	4.7e-06

似然比检验统计量=27.3, 自由度=3,  $P=5.06e-06$ ,  $n=193$ 。

表2 转移的系数回归结果

项目	Coef	exp(coef)	se(coef)	Z	P值
淋巴结	1.473	4.364	0.499	2.95	0.0032
肿瘤大小	0.374	1.454	0.338	1.11	0.2700
个人乳腺疾病史	-0.849	0.428	0.400	-2.12	0.0340

似然比检验统计量=16.9, 自由度=3,  $P=0.00073$ ,  $n=193$ 。

以上的R程序所做的是把淋巴结、肿瘤大小、个人乳腺疾病史作为影响复发和转移的协变量, 当  $\text{status}=1$  时关心的终点事件是复发, 当  $\text{status}=2$  时关心的终点事件是转移。从前一部分结果可以看出术前淋巴结的回归系数为0.23,  $P=0.47$ , 差异没有统计学意义。肿瘤大小的回归系数为0.860, 说明肿瘤越大复发的危险率越高。个人乳腺疾病的回归系数为-1.512, 说明患过乳腺疾病可以作为复发的一个保护因素, 没有个人疾病史的患者复发的危险率是有疾病史患者的4.53倍。第二部分的结果呈现了这3个因素对转移的影响。其中只有术前淋巴结的  $P$  值是有意义的, 回归系数是1.473。个人乳腺疾病史的回归系数为-0.849,  $P$  值为0.034, 差异有统计学意义。肿瘤大小的  $P$  值均不显著。以上结果说明同一个因素对不同的终点事件的影响是不同的。

在考虑竞争风险的情况下, 通过逐步筛选协变量, 得到肿瘤大小和个人乳腺疾病史是影响复发的独立预后因素。结果如表3所示。

### 2.2 非参方法的结果

运用非参方法得出的结果如表4所示。

表3 复发的独立预后因素分析

项目	Coef	exp(coef)	se(coef)	Z	P值	exp(-coef)
肿瘤大小	0.864	2.3726	0.2660	3.248	0.00116	0.4215
个人乳腺疾病史	-1.5811	0.2057	0.3362	-4.703	2.57e-06	4.8603

表4 复发和转移的累积风险率非参估计步骤

时间 $t_j$ (月)	风险集 $n_j$	总体生存 率 $\hat{S}(t_j)$	复发				转移			
			发生事件 数 $d_{1j}$	瞬间风险 率 $\hat{\lambda}_2(t_j)$	非条件风险 率 $\hat{p}_2(t_j)$	累积风险 率 $\hat{I}_2(t_j)$	发生事件 数 $\hat{d}_{2j}$	瞬间风险 率 $\hat{\lambda}_2(t_j)$	非条件风险 率 $\hat{p}_2(t_j)$	累积风险 率 $\hat{I}_2(t_j)$
1.4	213	0.9953	0	0	0	0	1	0.0047	0.0047	0.0047
2.93	203	0.9904	0	0	0	0	0	0	0	0.0047
3	202	0.9904	1	0.0050	0.0049	0.0049	0	0	0	0.0047
3.87	195	0.9904	0	0	0	0.0049	0	0	0	0.0047
4	194	0.9853	1	0.0052	0.0051	0.0100	0	0	0	0.0047
4.02	193	0.9802	1	0.0052	0.0051	0.0151	0	0	0	0.0047
4.06	192	0.9751	0	0	0	0.0151	1	0.0052	0.0051	0.0098
4.07	191	0.9700	0	0	0	0.0151	1	0.0052	0.0051	0.0149
5.03	184	0.9647	0	0	0	0.0151	1	0.0054	0.0053	0.0202
5	183	0.9647	0	0	0	0.0151	0	0	0	0.0202
5.1	182	0.9594	1	0.0055	0.0053	0.0204	0	0	0.0202	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
111.63	3	0.3554	0	0	0	0.3771	0	0	0	0.2675
124.4	2	0.1777	1	0.5000	0.1777	0.5548	0	0	0	0.2675
158.27	1	0.0000	1	1.0000	0.1777	0.7325	0	0	0	0.2675

有部分时间点的步骤省略。

表4显示了估计复发和转移累积风险率的步骤。例如，在 $t=1.4$ 月时，有1个患者发生了转移，此时风险集的病例数为213，瞬间风险率为 $1/213=0.0047$ ，由于在时间1.4以前既没有发生复发也没有发生转移，所以生存率为1，从而转移的无条件风险率为 $1 \times 0.0047=0.0047$ ，累积风险率为0.0047。

2.3 K-M方法与竞争风险模型的比较

生存分析得出的风险率曲线要高于竞争风险得出的风险率曲线，由K-M方法计算出的3年、5年复发累积风险率高于竞争风险计算出的结果，详见图1、表5。所以，如果不考虑竞争风险会造成累积风险率的高估，得出的结论与我们的理论推断一致。

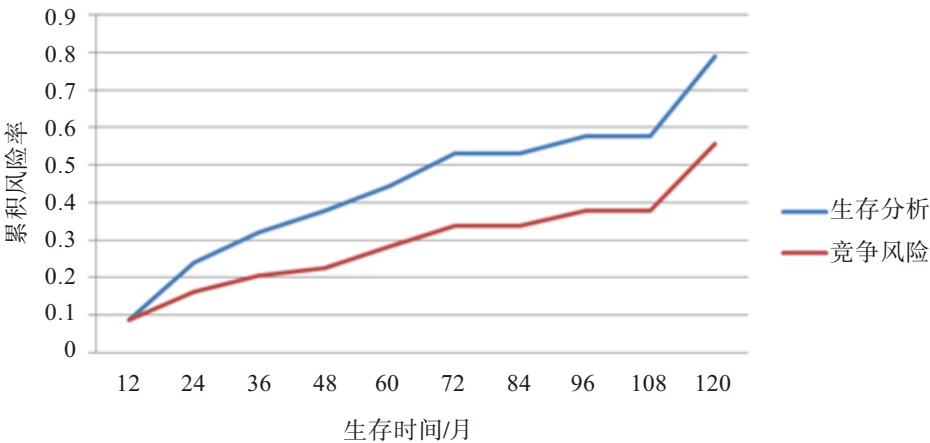


图1 两种不同分析方法的复发累积风险率比较

表5 K-M方法与竞争风险的非参方法估算的风险率比较

方法	时间/年	复发风险率/%	转移风险率/%
K-M方法	3	32	25.3
	5	44.4	36.5
竞争风险	3	20.73	17.2
	5	28.2	24.95

### 3 讨论

在考虑竞争风险的情况下,运用原因别比例风险模型对复发进行多因素分析,得到肿瘤大小和个人乳腺疾病史是影响复发的独立预后因素,淋巴结被排除了,但是术前淋巴结是影响转移的预后因素。肿瘤大小严重影响着乳腺癌患者的死亡和复发,肿瘤越大,患者的死亡和复发风险率越高,所以要早预防、早治疗来控制肿瘤的大小,另外应争取最大程度的肿瘤细胞减灭术才能真正提高乳腺癌患者的预后治疗效果。乳腺癌改良根治术相对于乳腺癌根治术的预后效果要更好。结果显示有个人乳腺疾病史的患者反而生存率更高,说明经常进行体检,关注身体的变化,进行有效的预防措施,会及早发现病症,及早诊断,及早治疗,从而使预后治疗效果更好。

由非参估计法计算复发和转移累积风险率,结果为:复发的3年风险率为20.73%,5年风险率为28.2%;转移的3年风险率为17.2%,5年风险率为24.95%。将非参方法的估算结果与K-M法估算的结果进行了比较,K-M方法的风险率高于竞争风险情况下的风险率,从某种程度上说明了K-M方法高估了风险率,符合前面理论部分的推断。

从方法上来说,国内现有的研究仅限于生存分析的应用上,很少有运用竞争风险模型来分析乳腺癌的预后研究。本文做了这方面的尝试,它能够更加精确、合理地分析预后因素对乳腺癌各个结局的

影响。此外,运用累积风险模型可以直观的了解终点事件的累积风险率<sup>[3-5]</sup>,近年来在国外的研究中此模型得到了推广,但国内的统计研究者几乎没有用到此模型。竞争风险的分析在一般的统计软件上没有相应的模块,需要用到R2.10.1里的cmprsk程序包来编程进行分析<sup>[6]</sup>。但是竞争风险模型也存在着缺陷,因为它只考虑最先发生的终点事件,忽略了后面的状态,如果要考虑后面整个预后过程,需要用到多结局模型,竞争风险模型仅是多结局模型的一种特殊情况,在以后的研究中可以尝试多结局模型。

从数据上来说,本文的数据是从福建厦门市4家医院现有的病例数据搜集而来的,存在着选择偏倚、失访偏倚以及混杂偏倚等,所以估计的结果不一定全面、准确。关于更加精确的结果有待于进一步扩大样本量、收集更广泛的资料并完善各个指标值。

### 参考文献:

- [1] 张敏璐,黄哲宙,郑莹. 中国2008年女性乳腺癌发病、死亡和患病情况的估计及预测[J]. 中华流行病学杂志, 2012, 33(10): 1049-1051.
- [2] 方积乾. 卫生统计学[M]. 北京: 人民卫生出版社, 2006: 369-374.
- [3] Putter H, Fiocco M, Geskus R B. Tutorial in biostatistics: Competing risks and multi-state model[J]. Statistics in Medicine, 2007, 26(11): 2389-2430.
- [4] 卢梓航,周立志,韩栋,等. 竞争风险型数据的统计处理及应用[J]. 现代预防医学, 2013, 40(5): 804-807.
- [5] 郭胜楠,陈慧林,李丽贤,等. 应用竞争风险模型探索宫颈癌患者预后的影响因素[J]. 肿瘤, 2015, 35(9):1006-1012.
- [6] R Development Core Team. R: a language and environment for statistical computing[EB/OL]. [2015-10-29]. <http://www.gbif.org/resource/81287>.